2012

Prospects for Scalability: Relationships and Uncertainty in Responsive Regulation

Cristie Ford

Allard School of Law at the University of British Columbia, ford@allard.ubc.ca

Citation Details

Cristie Ford, "Prospects for Scalability: Relationships and Uncertainty in Responsive Regulation" ([forthcoming in 2013]) 7 Reg & Governance 14.

Prospects for Scalability: Relationships and Uncertainty in Responsive Regulation

Cristie Ford

ford@law.ubc.ca

Ian Ayres's and John Braithwaite's book, *Responsive Regulation: Transcending the Deregulation Debate* (1992) was a significant step relative to the work that preceded it, and simultaneously reflective and emblematic of a larger contemporaneous shift. The book represents a convergence between rational choice / game theory analysis, which was perhaps then at its apogée, and the more sociological account being developed by people like John Braithwaite and Peter Grabosky (Gunningham and Grabosky 1998). Mary Ann Glendon's *Rights Talk* (1991) and Robert Ellickson's *Order Without Law* (1991) had both come out the year before. David Osborne's and Ted Gaebler's book *Reinventing Government* had come out the month before, in February 1992 – the same month that the Maastricht Treaty, creating the European Union, was signed. All in 1992, Yugoslavia fell apart, apartheid was voted away in South Africa, the UN held the Earth Summit in Brazil and, in November, Bill Clinton was elected President of the United States. These are significant markers on the regulatory timeline because many of the prominent works, and the prominent contexts, that we refer to today – the Open Method of Coordination in the EU, the Reinventing Government efforts of Al Gore and Bill Clinton, significant efforts toward global environmental law and regulation, even the novel extrajudicial concept of the Truth and Reconciliation commission – had not yet occurred. This is to say nothing of the revolutions we have seen since. The internet was in its infancy. The multiple preconditions to the recent financial crisis – including widespread complex financial engineering in general and consumer debt securitization in particular, and the extraordinary growth in the over-the-counter derivatives market – were not yet in place.

This paper seeks to take the original 1992 version of responsive regulation theory, as articulated by Ian Ayres and John Braithwaite in their book of that name, on its own terms. The significant insights *Responsive Regulation* gave us have transcended the book's time at least as much as those from any other contemporary model. Consider the enforcement pyramid, the benign big gun, tripartism, and the way in which a tit-for-tat regulatory stance is supported by both game theory and civic republican sociological approaches. *Responsive Regulation* developed a principled way for regulators to choose between punishment and persuasion, recognizing that neither approach works all the time. Consider also the concepts of "multiple selves" and multiple registers of action; the relationship that *Responsive Regulation* describes between regulation and other forces, be they public interest groups (PIGs) or other actors contributing to the "license to operate" (Gunningham et. al. 2003); and above all the embrace of contestability and revisability. *Responsive*

Regulation's insistence on pragmatic, context-sensitive application and continual evolution is an important commitment given the complexity and uncertainty that characterizes so many regulatory environments today.

Responsive Regulation develops a theoretical model that is meant to apply to a broad range of contexts. It makes a large claim, even while recognizing that the significance of the theory can be assessed only "through praxis in concrete institutional arenas" (Ayres and Braithwaite 1992, p. 99). In that spirit, our recent shared experience of the financial crisis may illuminate aspects of *Responsive Regulation* that might otherwise have gone unremarked. Twenty years on and with the benefit of this hard experience, two aspects of *Responsive Regulation* are striking. The first is the direct, personal relationship on which the regulatory interaction is premised. The second is the boundedness and manageability of the regulatory project. Regulators are understood to know what behaviors constitute compliance and non-compliance, to be able to interpret accurately signals from industry, and to be able to calibrate appropriate responses. At least in prudential regulation of global financial institutions in the wake of the recent financial crisis (though surely elsewhere too), neither the ongoing face-to-face relationship nor the boundedness or knowability of the regulatory terrain can be taken for granted.

This brief essay seeks to open a preliminary conversation about *Responsive Regulation* in terms of its *scalability*. It considers whether as a practical matter, *Responsive Regulation* can be scaled up to more diffuse, multiparty, logistically complex contexts, such as financial regulation. While the theory aspires to general applicability, it is grounded in empirical work in a particular kind of regulatory environment, meaning that it may be less applicable in others. As a matter of representation, the essay asks whether by projecting the focal object, the responsive relationship, outward, *Responsive Regulation* distorts our image of regulation in other contexts. In doing so, the inquiry inevitably reflects back on *Responsive Regulation*'s own home environment, where the question is whether *Responsive Regulation* also oversimplifies the complexity and challenges inherent in the interpersonal relationship itself. The essay closes by arguing that in order to incorporate responsive regulation's considerable discursive and relational benefits into regulatory environments such as global financial regulation, it needs to be buttressed by additional regulatory technologies.[i]

The Roots of Responsive Regulation

Responsive Regulation puts forward a personal, and interpersonally-based, perspective. Some of the book's deepest roots, on John Braithwaite's side, are with subject areas such as coal mine safety, or patient care in the nursing home industry. Accounts and interview excerpts derived from these environments give the book much of its narrative force. These are environments characterized by direct, face-to-face interactions between regulators and regulated actors. The fact that industry actors must be licensed *ab initio* generates an essentially finite and knowable pool of participants. Regulated actors are also geographically

embedded, in spaces such as coal mines and nursing homes, the physical condition of which is one of the things regulators are concerned about. The way to examine such institutions is to physically visit them. The context that orients *Responsive Regulation* is the personal nexus between compliance officer, supervisor, enforcement staffer, and management and workers.

In addition to being personal and face-to-face, responsive regulation is *interpersonal*, meaning that the quality and nature of the human relationship undergirding regulation is of primary concern. The enforcement and regulatory pyramids are institutional features built outward from a series of human interactions over time, not from organization-level design concerns. The orientation toward the interpersonal has only increased over the last twenty years. The conceptual link between the responsive regulatory ideal and John Braithwaite's restorative justice work (e.g., Braithwaite 1989), which is also rooted in personal relationships (e.g., Braithwaite 2002), is increasingly strong and central.

The interpersonal orientation makes responsive regulation a rich perspective in a field, regulation, which tends sometimes to revolve around more theoretical or structural accounts. The level of so-called mundane human engagement has great positive and normative significance, of course. As behavioral psychologists and others have forcefully demonstrated in the years since *Responsive Regulation* was published, we ignore the interpersonal at our peril.[ii] The interpersonal orientation is also the source of the book's great dynamism and context-sensitivity. From it flow its important insights around tailoring regulation to particular actors, accessing situational knowledge, and permitting flexible regulation through the mechanism of enforced self-regulation (Ayres and Braithwaite 1992, p. 101-132). A focus on the interpersonal makes it possible for *Responsive Regulation* to envision a specific new form of flexible regulation, distinct from both outright deregulation and from the increasingly maladaptive command-and-control bureaucratic technique that (conventional wisdom tells us) preceded the book.[iii]

This is not to say that *Responsive Regulation* is indifferent to more systemic questions. The book's civic republican commitments are clear throughout. The book anticipates subsequent work on the value of incremental problem-solving as a mechanism for regulatory policy-making.[iv] Ayres and Braithwaite envision an incremental accretion of well-designed and participation-enhancing regulatory moments – such as giving real powers to relevant PIGs within tripartite discursive structures – ultimately to be constitutive of a republican political order that emphasizes direct participation in the regulatory moment, often at a very local level, in the service of citizen empowerment and a thicker, more engaging account of citizenship (Ayres and Braithwaite 1992, p. 17-18).

In thinking about what *Responsive Regulation* can contribute to the financial regulatory arena, the question may be the degree to which one of its essential elements, the direct interpersonal relationship as a basis of knowledge and action, is scalable to the kinds of contexts that concern us in financial regulation.

Regulation and Scale

What is scalability as a functional matter? This essay uses the term scalability, first, in the sense that computer systems designers use it: as a technical measure of how well systems can handle increasing workload and data volume, either based on existing resources or by applying cost-effective strategies for extending the system's capacity (Weinstock and Goodenough 2006). A provisional definition of scalability in regulation would be that it is a measure of whether and how well a regulatory strategy operates in environments characterized by greater levels of logistical complexity, workload, and scope. Of course, regulation inevitably takes place in a multi-scalar and trans-scalar manner, and both regulation and relevant scale definitions shift with time and context (see, e.g., Osofsky 2009). Scale levels also affect each other (Ostrom, 2009).

All regulatory environments are characterized by complexity at the level of interpersonal relations – *Responsive Regulation* offers quite stylized depictions of them, even in its "home contexts" – but certain fields have other things going on besides. More complex, technical, or contingent regulatory contexts impose non-trivial loads on regulatory architecture. Some regulatory fields simply have to engage with more, and more far-flung, regulatees than others. Some regulatees operate across regulatory jurisdictions, national or otherwise. Some regulatees produce more products, and more complex products, more quickly. Some products are held to more technical or complicated regulatory standards than others.

This essay describes scalability through an example from international prudential regulation. This is by no means the only possible example of a logistically scaled-up regulatory context. But at least in the context of international financial regulation, the scope, volume, and logistical complexity of the regulatory task (meaning, e.g., the dynamic and highly technical nature of the subject matter, the "long tail" and uncertain nature of consequences, difficulties in developing outcome indicators or assessing compliance) make things like tit-for-tat engagement and compliance evaluations far more challenging than responsive regulation imagines. Highly diffuse or decentralized environments, like some of the transnational network variety, also increase complexity (Abbott and Snidal 2012). Scale is a function of a model, not the actors that implement it, though we could also encounter separate, serious problems of disarticulation or lack of coordination between actors at different scales (Abbott and Snidal 2012; Heimer 2011).

The concept of scalability assumes that there are different scales that are of concern to us. This much seems clear, although we could define scale in terms of any number of metrics including the geographic, organizational, technical, or product-oriented.[v] Because the core operative mechanism in *Responsive Regulation* is the direct, interpersonal relationship between regulator and regulated entity, the relevant technical question is whether and how well responsive regulation

functions when the relationships in question move from one-to-one, to many-to-many.[vi] The kind of scaling up we are concerned with is scaling up from a context where an inspector is engaged in a direct relationship with an inspected party, in relation to a bounded physical space, around a relatively straightforward set of regulatory compliance criteria – to contexts characterized by multiparty, attenuated, or disintermediated relationships, a larger and perhaps less clearly delineated regulatory space (in terms of scope, but also in the move from enforcement/compliance to prospective rulemaking), and a more logistically complex or contested set of regulatory compliance criteria.

In addition to being a technical measure, scale is also a conceptual. In a classic article from 25 years ago, Boaventura de Sousa Santos (1987) defined a conception of scale in law, by drawing an analogy between cartography and law. He claimed that "the relations law entertains with social reality are much similar to those between maps and spatial reality. Indeed, laws are maps …" (de Sousa Santos 1987, p. 282). Because maps cannot represent all features of the real world with perfect accuracy or they would have to be the size and shape of the real world, maps distort reality through three mechanisms: scale, projection, and symbolisation. This essay is most concerned with the first two.[vii] Speaking representationally, as de Sousa Santos does, *scale* is a function of size relative to the phenomenon to be mapped. As scale increases, that is the model tries to encompass a larger area, the representation of detail must decrease.[viii] Because maps are "a miniaturized version of reality, map-making involves the filtering of details, the selection of both meaningful details and relevant features" (de Sousa Santos 1987, p. 283). Whether a map is useful for a particular purpose depends on whether the details that have been selected are appropriate for those purposes.

De Sousa Santos describes *projection* in map-making as referring to the distortions that have to be built in, in order to represent a round globe on a flat piece of paper. Cartographers can choose different ways to distort the globe. They can trade off overlarge poles and undersized equators, or choose to depict distance accurately at the expense of depicting area accurately. Which compromises are most acceptable will depend on the purpose to which the map is put. Crucially for our purposes, projection and distortion also happen outward from a focal object or concern. As de Sousa Santos (1987, p. 285) says, "each map, each historical period or each cultural tradition of map-making has a centre, a fixed point, a physic or symbolic space in a privileged position around which the diversity, the direction, and the meaning of other spaces is organized". Whether the distortion of reality that inevitably accompanies cartography does or does not entail a "distortion of truth", as de Sousa Santos (1987, p. 282) puts it, is a function of the means by which scale, projection, and symbolisation are deployed.

For de Sousa Santos, the hermeneutic tools of scale and projection are useful for understanding law. Using examples from his empirical work around "revolutionary legality" in Portugal, property rights in Brazil, and popular courts in the Cape Verde Islands, de Sousa Santos claims that particular accounts are only coherent at particular scales. Although different scale level accounts interact,

and particular social events can amount simultaneously to "legal events" at more than one scale level, the legal accounts created at the different levels – in terms of motivations, actors, central issues – are not equivalent.

The same point can be made in financial regulation. At the small scale regulators could be concerned with investigating a particular financial institution for subprime mortgage fraud (Freiden 2004). This is the kind of factual scenario in which the *Responsive Regulation* account makes intuitive sense, even if regulators are not inspecting a physical plant. They are inspecting physical documents and speaking to those involved in creating them. Moving to a higher scale, regulators could be concerned with the various steps through which a defined group of financial institutions in a particular country transformed those subprime mortgages into triple-A rated securities, and marketed them. Moving to a still higher scale, regulators are confronted with the relationships between the marketing of those instruments and effects such as global systemic risk, system effects, interconnectedness, and the magnitude and impact of the over-the-counter derivatives market (see, e.g., Schwarcz 2008; McCoy et. al. 2009). Similarly, the BP Deepwater Horizon oil spill in summer 2010 produced, at different scales, lawsuits by families of those killed on the well; questions about safety in the oversight of deepwater drilling rigs; and extensive environmental effects (Broder 2011).

Extrapolating from de Sousa Santos, we might say that some conceptual constructs are defined (explicitly or implicitly) in terms of a particular scale, and only come into focus at that scale. Examples would include ecosystem-based environmental law, or the tragedy of the commons – a phenomenon that can only affect those that have a commons in common. Certain phenomena – a the law of large numbers in statistics, or herd immunity in epidemiology – are so contingent on achieving a particular scale level that their necessary preconditions are not present at other (in these examples, lower) levels. James Madison was making a scale argument when he said that freedom from oppression required a republic of a certain minimum size.[ix] Systemic risk, system effects, and groupthink also depend to some degree on achieving a particular scale. Where they are present, they exert a considerable effect. Models embedded within other scales, which cannot "see" system effects, will miss an important feature of the regulatory landscape.

Similarly, if the truth of an assertion depends on a scale precondition, then claims that seem to make sense at one level may seem incoherent at another. Lawrence Cunningham (2007) critiques the "rhetoric" of principles-based securities regulation on this basis. He argues that individual statutory provisions may be more rules-based or principles-based, but that the language is meaningless when used to describe entire regulatory regimes (but see Ford 2010, p. 265). Technologies and methods can also be tightly linked to a particular scale level. Apprenticeship-based education systems require one-on-one mentoring, while modern classroom education can occur in 500-seat lecture theatres. Absent some boosting technology, apprenticeships are not as scalable as lectures.

Closer to home (i.e., to *Responsive Regulation*), the same may apply to game theoretic modelling. A tit-for-tat model relies on a particularly direct and responsive regulator/regulated relationship. Each of two parties makes a single move, which is followed immediately by the other party's responding move. This is a highly stylized representation of regulatory interactions in any environment – an important point, though beyond this paper's scope – but the point here is that the consequential and signaling functions on which the model depends are even harder to generate when one moves from a one-to-one relationship to more complex, "noisier" environments. These would include one-to-many relationships, many-to-many relationships, or contexts where one party takes more than one move during its turn, and so cannot interpret exactly what provoked the regulatory response it receives.

The various accounts generated through scale functions are neither operatively nor normatively neutral. Precisely how the legal accounts are drawn at different scale levels is deeply constitutive of the social event itself (de Sousa Santos 1987, p. 288). Moreover, like the focal concern on a map, the central legal concerns will ramify outward:

> "conceptualisations, interpretive styles and techniques as well as ideological configurations dominant at the centre tend to be taken out of their context in which they originate and exported to (and imposed upon) the periphery. They are then applied in the legal periphery with little attention to local regulatory needs, since such needs are always interpreted and satisfied from the point of view of the centre." (de Sousa Santos 1987, p. 292.)

Scale, then, is both quantitative *and* qualitative. Talking about the general scalability of regulation is too vague to be helpful. We may be concerned simply about whether a system is operationally scalable at all (and it may be that some elements of some models are not). Yet we will also be concerned about whether the model scales up, while preserving attributes we particularly care about – its efficiency, stability, transparency, flexibility, responsiveness, or democratic representativeness. Taking steps to boost the scalability of a particular aspect of a model may negatively affect the scalability of other aspects. Assessing the "fit" of a particular model in relation to the regulatory problem at hand requires that we make normative choices and establish priorities, *ex ante*.

As Elinor Ostrom has identified, different functions are best accomplished at different scales. In a world of finite resources, imperfect information, and self-interested action, no single method can be ideally effective across all scales (Ostrom 1990). Scale-inappropriate modeling will be at best ineffective, and at worst deeply damaging.[x] If we are serious about regulatory effectiveness, there is value in trying to understand at what scale particular models are oriented, what

the implications may be for their broader effectiveness, and what we might be able to do about it.

How would we describe the scale level at which the responsive regulatory account is principally situated? In terms of *projection*, what kinds and degrees of distortions might we confront as we move out from *Responsive Regulation*'s core preoccupations? Notwithstanding *Responsive Regulation*'s broader republican commitments, the face-to-face compliance moment is its touchstone. Following on John Braithwaite's earlier book on the subject (1987), the question of when to punish and when to persuade is also this book's central concern. As Ayres and Braithwaite recognize (1992, p. 58), their method is about enforcement and compliance, not prospective rule-making. To a degree that is quite unique in regulation studies, the circumstances are within the range of what Abram Chayes (1976) would have recognized as a fairly traditional conception of adjudication. This impresses the entire account with a particular orientation toward optimizing regulatory effectiveness in those moments. In de Sousa Santos's terms, that moment is the focal concern of the responsive regulatory map, and its conceptualizations and techniques are exported to the broader context. What *Responsive Regulation* offers, then, is a roadmap for how to navigate through regulatory interactions in a world where the players are depicted as known or knowable, the subject matter is understood to be discrete, and what is called for is a rational decision making process that regulators can follow in determining when to lean in on an industry actor, and when to ease off.

It must be said that we should consider whether even the "home environment" that *Responsive Regulation* addresses is necessarily as simple as it seems, or as simple as responsive regulation claims it to be. Leaving aside the behavioral psychological advances alluded to above, which substantially complicate our understanding of the interpersonal relationship, recent empirical work testing responsive regulation as implemented also suggests considerable gaps between theory and practice (Mascini and Van Wijk 2009; Nielsen and Parker 2009; Welsh 2009; also Etienne 2012). This short essay cannot engage with these challenges, which are beyond its scope. The point here is that, *even if* we assume a high degree of congruence between responsive regulatory theory and real life experience in the theory's home environments, we cannot assume that responsive regulation will automatically scale upward into the highly fluid, mutable, and dynamic scale at which global financial regulation operates. The example below seeks to illustrate this point. The final section of the essay considers options that may permit the nature of the knowledge generated in the responsive regulatory relationship scale up, even if the relationships themselves do not.

Basel II as a scale problem for enforced self-regulation

Relative to the command-and-control literature that preceded it, responsive regulation represents an intentional move toward a more sophisticated, multi-level regulatory approach. Responsive regulation recognizes that a lot of information is

best gathered in a decentralized, firm-level manner and that knowledge about the particular risks associated with a particular line of business should also be drawn upward to the regulatory level. Though Ayres and Braithwaite do not frame it in these terms, one could think of the enforcement pyramid as a compliance-into-regulation scaling tool, designed to locate discrete regulatory interactions within a broader conceptual matrix and to aggregate them into a comprehensive regulatory stance. Tripartism, also, implicitly recognizes that there are interested parties operating at the interstitial layer of "civil society" (as it then was called), who could be brought into the regulatory conversation. The book's account of partial industry regulation (Ayres and Braithwaite 1992, p. 133-157), that is, that government can effectively delegate regulatory tasks to a firm's competitors by forcing unregulated firms to compete with regulated ones, seems like a coverage-oriented scaling strategy, in the sense that it extends the effect of regulation beyond the directly regulated parties. *Responsive Regulation* also explicitly notes that both individuals and firms/corporations are capable of disaggregation into "multiple selves" (p. 30-35) – a vivid recognition of scale variability along a wholly different, actor-oriented parameter.

Note that, consistent with Responsive Regulation's focal concern on relationships, each of these ways of addressing scale is fundamentally *relational* in orientation. To the degree that *Responsive Regulation* even implicitly speaks about scale, it does so from the point of view of the interpersonal and relationship-based priorities that animate it. Indeed, responsive regulation as described by Ayres and Braithwaite in 1992 is optimized in terms of efficiency, reliability, and regulatory credibility when operated at smaller scales.

To illustrate, let us describe the June 2004 Basel II Capital Accord (Basel II) as analogous to enforced self-regulation in *Responsive Regulation*'s terms. I am not saying the financial crisis was caused by Responsive Regulation, or saying that Basel II was a faithful instantiation of enforced self-regulation. It was not. The financial crisis was multifactorial and other significant problems, such as regulatory gaps, played far larger roles. Nevertheless there is some value in using the financial crisis analogically, to help foreground the ways in which any regulatory approach is bound to a particular scale and ill-suited to environments where its focal objects are not the operative drivers.

Responsive Regulation locates "enforced self-regulation" at the crucial intermediate layer of the enforcement pyramid (p. 101-132). This is an arrangement under which firms develop their own set of context-specific conduct rules, which are then publicly ratified and capable of public enforcement. The Basel II Capital Accords were this kind of arrangement. Basel II establishes high-level, outcome-oriented requirements around the amount of capital that financial institutions need to maintain in reserve, and then devolves the process-based risk assessment details to the institutions themselves. The greater risk a financial institution was carrying, the greater its reserves had to be. In 2004, Basel II capital adequacy formulae were also incorporated, to disastrous effect, into the United States

Securities and Exchange Commission's Consolidated Supervised Entities program (CSE Program), concerning capital requirements for leading broker-dealers.

Let us now consider the assumptions and core preoccupations of enforced self-regulation. Fundamental to responsive regulatory strategies such as enforced self-regulation is, *first*, the assumption that regulated actors (here, "firms") are rational, autonomous and coherent enough to know how they are conducting themselves internally, and to behave in their own self-interest. Like other game theoretic models, tit-for-tat would make no sense if a regulator did not have a rational counterparty to deal with (Becker 1968). *Second*, enforced self-regulation assumes a meaningful regulatory presence. The regulator must be in a position to credibly verify firms' conduct and to respond accordingly. It should hold in reserve a "benign big gun" that it can be expected to deploy, predictably and with justification, where circumstances warrant. For Responsive Regulation, the source and the context of that knowledge is the direct, highly communicative, interpersonal relationship between regulator and firm. Responsive Regulation does not require that the regulator be in the best position to know all the details of a firm's business (quite the opposite), but it does require that the regulator actually have knowledge about the firm, including perhaps its "stance" (Black and Baldwin 2007). The regulator needs to be in a position to measure firms' conduct against broader regulatory standards. *Third*, Responsive Regulation is framed around the notion that the regulator/firm relationship is the primary locus where important things happen. This is the rationale behind the tripartism prescription. Tripartism assumes that the main way to give voice and effect to other, civil society priorities is to make them part of *that* conversation.

What actually took place around prudential regulation of global financial institutions under Basel II (and the CSE Program) was quite different, and in a sense can be understood as problems of scale. Finance today is global, electronic, fast-moving, and based on esoteric knowledge in a way that pulls it away significantly from regulation of physically embedded and static regulated actors. The first assumption above – that firms are rational, autonomous, and coherent enough to behave in their own self-interest – was disproved in part as a function of scale. Firms' capacity to know, internally, exactly what risks they were running was severely limited by their heavy reliance on software to handle the extraordinarily complex assets, products, and markets they were dealing with (Schwarcz 2009). The software was flawed (Taleb 2007; Gerding 2009). Just as problematically, through its design the software buried contestable assumptions below the level of human judgment, so that even the firms themselves actually knew less about the kinds of risks they were running (Gerding 2009; Bamberger 2010).

The second assumption above, that of a credible regulatory presence, was also disproved in part as a function of scale. Revolving door arguments aside, the recent financial crisis cannot be understood primarily in terms of interpersonal relationships.[xi] In the prudential regulation environment, as a factual matter there was simply not a bedrock of personal relationship or personal knowledge on

which to build (Kingsford-Smith 2011). Among some key prudential regulators, such as the United Kingdom's Financial Services Authority, the relationship between firm and regulator became severely attenuated as a function of size, staffing, and diffuse priorities (FSA Internal Audit Division 2008; also Black 2010, p. 18-19). While regulators recognized the need to pursue "the same PhD rocket scientists the banks [were] chasing" (Hughes 2008), as a practical matter they lacked the regulatory capacity to credibly verify bank capital adequacy. In the marketing of increasingly risky products and the increasing leverage that financial firms themselves took on, the reliance on private sector financial modelling was amplified by the delegation of crucial decision-making, by regulators, to those-that-modeled. Regulatory self-accounts relied increasingly on the fiction of the self-disciplining efficient market (UK Financial Services Authority 2009). This was a scale problem in the sense that the scale of banks' resources and the scale of the problem outmatched regulators' resources. (Power and influence are more than just scale problems, of course: Ford 2011).

There was also a structural scale problem that contributed: in the United States, where the worst problems occurred, the Basel II structure was incorporated into the CSE Program on a *voluntary* basis because no American regulator had the authority to impose capital adequacy requirements on global shadow banks (SEC 2008, p. 81).[xii] Regulatory arbitrage between London and New York also played a role (Turner Review 2009). These are problems of scale mismatch between the national mandate of the regulator and the global scale of the issue. The result, in responsive regulatory terms, was that there was no credible public regulatory presence at all, and no benign big gun in the background.

Finally, the third assumption above – that the regulator/firm relationship was the primary locus where important things happened – was also disproved in part as a function of scale. Far more important in the run-up to the financial crisis were fully unregulated nearby spaces, such as the over-the-counter derivatives market and the players in it; and the industry-wide competitive effects that produced a behavioral cascade toward excessive risk-taking. Focusing on the regulator/firm relationship misapprehends the economic and collective scale at which problems were building.

To the extent that *Responsive Regulation* is organized around the three assumptions above, and especially on the presence of an interpersonal regulator-firm relationship as a basis for knowledge, it is incapable of speaking to the circumstances surrounding Basel II. Because interpersonal and knowledge-generating relationships are at the centre of the enforced self-regulation conceptual map, it would be a costly mistake – a distortion of *truth*, in de Sousa Santos's terms (1987, p. 282) – to apply the model upward through scales without reflecting on whether it still adequately represented the most salient facts, and whether it was still congruent with regulatory concerns and priorities that came into view at a different scale. Making enforced self-regulation relevant and meaningful at a wholly different scale requires that we find ways to scale the model up, past the reach of interpersonal relationships.[xiii] Moreover, we will want

to do so without losing whatever it is that we think is most essential about responsive regulation itself. While distortions and trade-offs will be inevitable as we move through scales, we can make choices about what we most care about accurately projecting upward.

Where do we go from here? Responses

The precise kinds of relationships on which *Responsive Regulation* is founded probably cannot be scaled upward without incurring excessive costs. They would be if we committed to a direct supervisory relationship all the way up, through national and even global financial regulation (e.g., Pan 2011), though such linear scalability would be very costly. Another option would be a form of corporatist or "club" government, in which important regulatory decisions were made in forums where regulatory and industry elites interacted with each other, with only an attenuated relationship to those they notionally represented. Financial institutions' and governments' mutual economic dependence already makes this a reality at the global level (Gelpern 2011). This preserves the direct relationship, but at the cost of transparency, incorporating local information, and representativeness and accountability (to say nothing of *Responsive Regulation*'s more aspirational civic republican aims). Instead, I would take a normative stance in favor of trying to project upward the nature and source of *information* that responsive regulation generates – incremental, contextual, experiential, and collaboratively generated.

There are two main options in trying to make responsive regulation function at higher scales. The first is to reduce the demands imposed on the responsive regulatory system, by limiting the scale, scope, or logistical complexity of the environment in some way. Computer systems designers know that systems are more scalable if they do less for each user. What, then, is the minimum that the (responsive) regulatory system must do (in order to still be responsive)? Risk-based approaches, which try to allocate regulatory resources to the most high risk contexts, may be helpful in this regard (e.g., Black 2005; but see Gunningham 2011, p. 9). We may also want to consider the selective use of prophylactic or default rules (e.g., Dorf and Sabel 1998), for the purpose of containing complexity and limiting explicit variability.[xiv] Bright line capital adequacy requirements are such a tool, and form part of the new Basel III Accord in the forms of leverage ratios, mandatory capital conservation buffers, and similar measures (Basel III). Routinization and standardization can also be scaling mechanisms, because by holding some elements stable they create a platform on which diverse approaches can interact (Simon 2011). Ironically, it was in fact a standardized contract for purchase and sale, the ISDA Master Agreement, which permitted the over-the-counter derivatives market to grow to the extent and in the diversity that it did (Jomadar 2007). Where and how risk assessments, standardization, and prophylactic rules are embedded is a value-laden choice. Additionally, as Annelise Riles has observed in describing the "agency" of tools, forms and technique are contested and highly consequential (2011, p. 229). Yet carefully applied, these

techniques may play an important role in corralling the regulatory project within manageable bounds.

The second option would be to try to boost responsive regulation's capacity at higher scales. We have fewer real-life examples to look to here, though scholars have contributed proposals. For example, creating a centralized standard-setting and information-processing clearinghouse to aggregate information and coordinate more localized regulatory engagements could help transform small-scale, discrete regulatory moments into something more systemic (e.g., Dorf and Sabel 1998, p. 287–89, 354–56). Institutionalized "learning loops" and the meta-regulatory notion of "regulating self-regulation" are another possibility (Parker 2002). Automated information gathering and computerized analytical models are surely part of this endeavor. Recognizing that the risks associated with automation can be considerable (Gerding 2009; Bamberger 2010), scalability to extremely complex environments would otherwise be beyond human capacity. Consciously establishing or trying to tweak existing governance nodes is another possibility, though we should be clear that there is nothing automatically democratic or fair about nodal governance (Burris, Drahos, and Shearing 2005).

In trying to scale responsive regulation upward, we cannot assume that either its dynamism or its representativeness will automatically flow upward. Elinor Ostrom's work is instructive here. She and her colleagues argue that workable regulation must reflect the boundaries of the relevant epistemic communities, and the problem in question. Efforts at scaling should pay attention to those boundaries, and should ensure that local-level information is valued and that agency relationships have some legitimacy (McGinnis and Ostrom 2008). Ayres's and Braithwaite's role for PIGs in tripartism reflects the same intuitions, but *Responsive Regulation* does not address exactly how PIGs will be constituted, and it does not contemplate a public role in ensuring that all relevant voices are at the table. What Ostrom (1990) envisions is much more intentionality in the design of *nested* regulatory enterprises across scales for such functions as information-gathering, policy-making, monitoring, enforcement, conflict resolution, and governance.

The trade-offs involved in scaling up *Responsive Regulation* are not unmitigated tragedies. As de Sousa Santos points out, representations can actually be made *more* useful if one includes and omits the right things as scale increases and detail is lost (1987, p. 283-84). The techniques above could make the difference between an effective regulatory process that builds on incremental engagements, and a series of disjointed and non-aggregating enforcement interactions that cumulatively do not a system make. The task is to identify the tools that can help convert diffuse, face-to-face responsive regulatory interactions into something both effective at a higher scale, and also reflective of *Responsive Regulation*'s own normative commitments in favor of civic engagement.

References

Legislation and Government Documents

Alternative Net Capital Requirements for Broker-Dealers that Are Part of Consolidated Supervised Entities, Exchange Act Release No. 34-49830, 69 Fed. Reg. 34,428 (June 21, 2004), available at http://www.sec.gov/rules/fmal/34-49830.pdf

Basel Comm. on Banking Supervision, Bank for International Settlements (2004) International Convergence of Capital Measurement and Capital Standards: A Revised Framework.

Basel Comm. on Banking Supervision, Bank for International Settlements (2010-2011) International Convergence of Capital Measurement and Capital Standards: A Revised Framework.

FSA Internal Audit Division (2008) The Supervision of Northern Rock: A Lessons Learned Review, online: Financial Services Authority <http://www.fsa.gov.uk>

SEC's Oversight of Bear Stearns and Related Entities: The Consolidated Supervised Entity Program (2008) Report No 446-A, online: Securities and Exchange Commission <http://www.sec.gov>.

UK Financial Services Authority (2009) The Turner Review: A Regulatory Response to the Global Banking Crisis, online: Financial Services Authority <http://www.fsa.gov.uk>

Secondary Sources

Abbott KW, Snidal D (2012) Taking Responsive Regulation Transnational. __ *Reg & Gov* __ (2012)

Ayres I, Braithwaite J (1992) *Responsive Regulation: Transcending the Regulatory Debate.* Oxford University Press, New York, NY.

Bamberger K (2010) Technologies of Compliance: Risk and Regulation in a Digital Age. *Texas Law Rev.*, 88, 669.

Becker G (1968) Crime and Punishment. *The Journal of Political Economy,* 76, 169–217.

Black J (2005) The development of risk-based regulation in financial services: just 'modelling through'?. In: Black J, Lodge M, Thatcher T (eds.), *Regulatory Innovation: A Comparative Analysis*, Edward Elgar, Cheltenham, UK.

Black J (2010) The Rise, Fall and Fate of Principles Based Regulation. LSE Law Society and Economy Working Papers 17, online: London School of Economics <http://eprints.lse.ac.uk>

Black J, Baldwin R (2007) Really Responsive Regulation. LSE Legal Studies Working Paper No. 15.

Braithwaite J (1989) *To Punish or Persuade.* State Uni. New York Press, New York.

Braithwaite J (1989) *Crime, Shame, and Reintegration*. Cambridge University Press.

Braithwaite J (2002) *Restorative Justice and Responsive Regulation*. Oxford University Press.

Braithwaite J (2009) Restorative Justice for Banks Through Negative Licensing. *Journal of Criminology*, 49, 439-450.

Breyer S (1982) *Regulation and Its Reform*. Harvard University Press.

Broder JM (2011) BP Shortcuts Led to the Gulf Oil Spill, Report Says. *New York Times*, Sept 14, 2011.

Burris SC, Drahos P, Shearing CD (2005) Nodal Governance. *Australian Journal of Legal Philosophy*, 30.

Chayes A (1976) The Role of the Judge in Public Law Litigation. *Harv. L. Rev.*, 89, 1281.

Chuang J (2001) Economics of Scalable Network Services. *Proceedings of SPIE ITCOM 2001*.

Cunningham LA (2007) A Prescription to Retire the Rhetoric of "Principles-Based Systems" in Corporate Law, Securities Regulation and Accounting. *Vanderbilt Law Rev.,* 60.

de Sousa Santos B (1987) A Map of Misreading: Toward a Postmodern Conception of Law. *Journal of Law & Society, 14,* 279-302.

Dorf M, Sabel C (1998) A Constitution of Democratic Experimentalism. *Colum L Rev*, 98, 267-473.

Duboc L, Rosenblum D, Wicks T (2007) A Framework for Characterization and Analysis of Software System Scalability. *Proc. ESEC-FSE'07*, 375–384.

Ellickson R (1991) *Order without Law: How Neighbors Settle Disputes*. Harvard University Press.

Etienne J (2012) Ambiguity and Relational Signals in Regulator-Regulatee Relationships. __ *Reg & Gov* __ (2012)

Ford C (2010) Principles Based Securities Regulation in the wake of the financial crisis. *McGill Law J.*, 55.

Ford C (2011) Macro- and micro-level effects on responsive financial regulation. *UBC L Rev* 44, 589-626.

Freiden T (2004) FBI Warns of Mortgage Fraud Epidemic. *CNN,* Sept. 17, 2004.

Gerding EF (2009) Code, Crash, and Open Source: The Outsourcing of Financial Regulation to Risk Models and the Global Financial Crisis. *Washington Law Review.*

Glendon MA (1991) *Rights Talk.* The Free Press, New York, NY.

Gunningham N (2011) Assessing Responsive Regulation 'on the ground': Where does it work? Paper presented at the annual meeting of the Law and Society Association, San Francisco, CA.

Gunningham N, Grabosky P (1998) *Smart Regulation: Designing Environmental Policy*. Clarendon Press, Oxford, UK.

Gunningham N, Kagan R, Thornton D (2003) *Shades of Green: Business, Regulation, and Environment.* Stanford Law and Politics, Stanford, CA.

Heimer C (2011) Disarticulated Responsiveness: The Theory and Practice of Responsive Regulation in Multi-Layered Systems. *UBC L Rev* 44, 663-694.

Hughes J (2008) FSA admits failings over Northern Rock. *The Financial Times*, online: Financial Times <http://www.ft.com/cms/s/0/0833a416-fb0d-11dc-8c3e-000077b07658.html?nclick_check=1>.

Jomadar B (2007) The ISDA Master Agreement - The Rise and Fall of a Major Financial Instrument. Online: http://ssrn.com/abstract=1326520.

Kingsford-Smith D (2011) A Harder Nut to Crack? Responsive Regulation in the Financial Services Sector. *UBC L Rev* 44, 695-.

Lindblom (1959) The science of muddling through. *Public Administration Review*, 19(2), 79-88.

Madison J (2003) Federalist 10. In: Rossiter C (ed) *The Federalist Papers*, pp. 71-79. New American Library, New York, NY.

Madison J (2003) Federalist 55. In: Rossiter C (ed) *The Federalist Papers*, pp. 338-342. New American Library, New York, NY.

Mascini P, Van Wijk E (2009) Responsive Regulation at the Dutch Food and Consumer Product Safety Authority: An Empirical Assessment of Assumptions Underlying the Theory. *Reg. & Gov.* 3, 27-47.

McGinnis M, Ostrom E (2008) Will Lessons from Small-Scale Social Dilemmas Scale Up? In: Biel A, Eek D, Gärling T, and Gustafsson M (eds), *New Issues and Paradigms in Research on Social Dilemmas,* pp. 189-211. Springer, New York, NY.

McCoy PA, Pavlov AD, Wachter SM (2009) Systemic Risk through Securitization: The Result of Deregulation and Regulatory Failure. *Conn. L. Rev.*, 41, 493-541.

Nielsen VL, Parker C (2009) Testing Responsive Regulation in Regulatory Enforcement. *Reg. & Gov.*, 3, 376-399.

Osborne D, Gaebler T (1992) *Reinventing government: how the entrepreneurial spirit is transforming the public sector*. Penguin Group, New York, NY.

Osofsky H (2009) Scaling 'local': the implications of greenhouse gas regulation in San Bernardino County. *Michigan Journal of International Law*, 30, 689-705.

Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

Ostrom E (2009) Design Principles of Robust Property Rights Institutions: What Have We Learned?. In: Ingram GK, Hong YH (eds) *Property Rights and Land Policies*, pp.25-51. Lincoln Institute of Land Policy, Cambridge, MA.

Pan E (2011) Understanding Financial Regulation. Cardozo Legal Studies Research Paper No. 329, online: http://ssrn.com/abstract=1805018.

Parker C (2002) *The Open Corporation: Effective Self-Regulation and Democracy*. Cambridge University Press.

Riles (2011) *Collateral Knowledge: Legal Reasoning in the Global Financial Markets*. Uni. of Chicago Press, Chicago, IL.

Schwarcz SL (2008) Systemic Risk. *Georgetown Law Journal,* 97(1).

Schwarcz SL (2009) Regulating Complexity in Financial Markets. *Washington Uni. Law Rev.*, 87, 211.

Scott JC (1999) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, New Haven, CT.

Short JL (2012) The Paranoid Style in Regulatory Reform. Georgetown Public Law Research Paper 11-10, online: http://ssrn.com/abstract=1739015.

Simon (2011) The Institutional Configuration of Deweyan Democracy. Columbia Public Law Research Paper No. 11-286; Stanford Public Law Working Paper No. 1957332, online: http://ssrn.com/abstract=1957332.

Sparrow M (2000) *The Regulatory Craft*: *Controlling Risks Solving Problems, and Managing Compliance*. Brookings Inst. Press.

Stewart RB (1981) Regulation, Innovation, and Administrative Law: A Conceptual Framework. *Calif. L. Rev.* 69, 1259.

Taleb NN (2007), *The Black Swan: The Impact of the Highly Improbable*, Random House.

Weinstock CB, Goodenough JB (2006) On System Scalability. Carnegie Mellon University Technical Note CMU/SEI-2006-TN-012, online: http://repository.cmu.edu/cgi/viewcontent.cgi?article=1400&context=sei

Welsh M (2009) Civil Penalties and Responsive Regulation: The Gap between Theory and Practice. *Melbourne U L Rev,* 33, 908-33.

---

[i] I am grateful to Carol Heimer for this phrasing.

[ii] Although not the focus of this essay, this is another significant way in which our understanding of regulation has evolved. When *Responsive Regulation* was published, Daniel Kahneman and Amos Tversky were still a full decade away from being awarded the Nobel Prize. Much behavioral psychology research engages at the same interpersonal level that *Responsive Regulation* does, and it represents an especially direct challenge to the book's more rational-actor based elements, including to some degree tit-for-tat enforcement and the regulatory pyramid. In a different vein, we also have greater insight now into the degree to which organizational hierarchies and legal architectural features influence the interpersonal/relational level at which *Responsive Regulation* is pitched.

[iii] An interesting recent empirical study of law scholars' increasing, and increasingly pejorative, use of the term "command-and-control" is in Short (2012).

[iv] See, e.g., Sparrow (2000); also the experimentalists, notably Dorf and Sabel (1998).

[v] De Sousa Santos (1987, p. 287) restricts his depiction of scale in law to three scale levels described geographically: local law, nation state law, and world legality. In other contexts, these three levels may seem more arbitrary than helpful. Another aspect of scale – the temporal dimension – is beyond this paper's scope. See, e.g., McGinnis and Ostrom (2008).

[vi] Scale as I am using it should be distinguished from the economic concept of "economies of scale". My question is not whether Responsive Regulation can enjoy increasing returns with scale (something I would seriously doubt) but rather whether the things we value about Responsive Regulation can be carried up through scale levels where the interpersonal relationship no longer pertains. On the distinction between economies of scale and scalability, see Chuang (2001).

[vii] Symbolisation, which de Sousa Santos (1987, p. 285) describes as "the representation of selected features and details of reality in graphic symbols", is less central to this discussion. The particular metaphors in Responsive Regulation – the enforcement pyramid, the benign big gun, the tit-for-tat relationship – are symbolisation.

[viii] Cartographers, and de Sousa Santos, use scale terms in the opposite way to how computer systems designers and some other academics use them. Cartography uses the term "large scale" to refer to more local, high-detail maps that cover a smaller area, and "small scale" for more high-level, less detailed maps. By contrast, computer systems designers and others speak of "scaling up" from "small scale" environments to more complex or larger ones. See, e.g., note 20 in McGinnis & Ostrom, 2008. Although the cartographic definition is more technically correct, I am using the computer science definition. It comports better with colloquial understandings of the terms and better illustrates the concepts examined here.

[ix] See, e.g., Federalist No. 10 (Madison 2003, p. 78); also Federalist No. 55 (Madison 2003, p. 339).

[x] See, e.g., Scott (1999).

[xi] This is not to dismiss those arguments; only to say that the primary mechanisms here were not directly interpersonal in the way that, for example, bribery or improper influence would be. The mechanism was not operating at the same level that Responsive Regulation would have been operating. It operated at the institutional level.

[xii] See also Ostrom (1990) on the need for congruence in terms of scale between regulatory mandate and regulatory capacity.

[xiii] John Braithwaite himself disagrees that this level of interpersonal distance is inevitable. He advocates instead for more hands-on regulation, which would allow regulators to "kick the tyres" on financial products. Braithwaite (2009).

[xiv] I am not saying that rules are invariably more certain or predictable than more flexible processes. Rigid rules may just force variability and discretion "underground". Nevertheless there is still some potential benefit in limiting the number of moving pieces with which a regulator must explicitly contend.